A Study of Effective Information for AI-Aided Medical Diagnostics

Alexander Han

Independent Self-Paced Research

Acton-Boxborough Regional High School

Acton, Massachusetts

08/22/2024

Abstract

This research evaluates the effectiveness of common parameters used in generative AI for medical diagnostics. These parameters include symptoms, medical history, medical tests, and medications. The study assesses effectiveness by providing detailed descriptions of these parameters to AI models GPT-3.5 and GPT-4o and analyzing their responses. Statistical analyses are conducted on the results. Findings indicate that symptoms and medical history are the most critical factors in medical diagnostics. While medical tests are essential for diagnosing certain conditions, they are irrelevant for others. Medications, however, do not significantly impact the diagnostic process.

Literature review

Chatbots have experienced a rapid surge in popularity, finding applications across various sectors globally. Among them, OpenAI's ChatGPT stands out as the most renowned. Launched on November 30th, 2022, it quickly captivated public interest. Although not the first of its kind, ChatGPT garnered significant attention due to its advanced capabilities compared to other chatbots. OpenAI leverages artificial intelligence (AI) and machine learning, enabling ChatGPT to provide accurate responses to inquiries [1], [2], [3], [4], and [5]. At that time, GPT-3.5 had 20 billion parameters, while GPT-3 had 175 billion parameters, enabling it to perform a wide range of complex tasks effectively [6], [7], and [8]. Having fewer parameters allows for quicker responses but limits the ability to handle more intricate tasks. Recently, OpenAI introduced GPT-4 and GPT-4o, with approximately 175 billion and 200 billion parameters, respectively. These models can perform much more complex functions than any previous chatbot, although at a slower pace than GPT-3.5 [9]. GPT-4o, the latest version, has seen significant improvements in areas such as accuracy, processing, and response speed. OpenAI has made GPT-4o accessible for free for the first few questions, and users can choose between GPT-3.5 and GPT-4o. Since the release of ChatGPT, many specialized chatbots have emerged in fields like finance, research, and medicine [10]. This study focuses on medical chatbots. In this paper, we refer to the development of an AI-based self-diagnostic tool as AI-Aided Medical Diagnostics (AAMD) due to the lack of formal terminology.

The development of AI-Aided Medical Diagnostics (AAMD) could greatly benefit individuals traditionally underserved by the healthcare system. In 2021, approximately 30 million Americans (9.2% of the U.S. population) lacked health insurance, often citing high costs and limited coverage as primary reasons. Hispanics were the minority group most likely to be uninsured, with 30.1% of Hispanic adults without health insurance. Consequently, these individuals often face higher medical costs and may avoid seeking medical care. A self-diagnosis tool would allow people to identify potential health issues, facilitating treatment while avoiding the high costs of professional diagnoses. Non-native English speakers could also benefit, as chatbots can communicate in various languages. Additionally, enabling patients to self-diagnose would enhance privacy regarding their symptoms and encourage more active participation in the diagnostic process, fostering a more open and collaborative relationship between patients and healthcare providers [11].

Numerous efforts are underway to integrate AI into the medical field. In 2017, Woebot, a chatbot designed to assist with mental health issues such as depression, anxiety, and addictions, was introduced to the public [12]. Woebot uniquely uses emojis to enhance nonverbal communication and better connect with users. Additionally, hospitals are increasingly utilizing AI for medical imaging, including CAT scans, MRIs, and X-rays, as AI can detect patterns or abnormalities that might be missed by the human eye [13]. Top U.S. hospitals like Mayo Clinic, Cleveland Clinic, Massachusetts General Hospital, Johns Hopkins Hospital, and UCLA Medical Center have incorporated AI into their programs. Mayo Clinic is working on using AI to provide more personalized treatment options for cancer patients [14], [15], and [16]. Cleveland Clinic employs AI to identify patients at high risk of cardiac arrest who need a vasopressor [17] and [18]. Massachusetts General Hospital uses its extensive collection of 10 billion medical images to train AI for radiology and pathology [19]. Johns Hopkins integrates AI into its command center to improve communication between medical teams and enhance ambulance dispatch, patient triage, and patient discharge processes [20] and [21]. UCLA Medical Center has deployed a chatbot called Virtual Interventional Radiologist (VIR) to help clinicians respond to common questions with evidence-based answers [22]. These AI integrations could significantly impact the future of hospitals, greatly increasing their efficiency and effectiveness [23].

Despite ongoing concerns about job displacement due to AI-Aided Medical Diagnostics (AAMD), most physicians worldwide are enthusiastic about the potential of AI as a diagnostic tool when used appropriately [24]. A survey of radiologists revealed that 89% were not worried about job loss, and 77% supported the integration of AI into radiology [25] and [26]. Many doctors are particularly excited about AI's potential to enhance the diagnostic process, boost efficiency, and improve clinical outcomes [27]. AI has already made significant progress in diagnosing various types of cancer through machine learning and natural learning capabilities, enabling doctors to provide more accurate treatments and potentially reducing cancer-related deaths significantly [28]. Meanwhile, Harvard Medical School is advancing AI education and implementation in healthcare by allowing students to use chatbots for diagnostic purposes [29].

Public opinion on AI-Aided Medical Diagnostics (AAMD) is nearly split, with 49% in favor and 51% against AI usage. Those opposed mainly cite concerns about privacy violations and a lack of understanding of how AI operates. Enhancing patient knowledge about AI could significantly improve its perception in the medical field, as 65% of patients indicated they would feel more comfortable if doctors explained how AI is used in medicine and healthcare [24].

To encourage the adoption of AI-Aided Medical Diagnostics (AAMD) technology, continuous efforts are needed to test and document its strengths and limitations for each specific disease. AAMD chatbots are more likely to misdiagnose rare diseases due to limited training data. Even GPT-4o struggles with rare disease diagnoses but is highly reliable for common illnesses like the flu or COVID-19 [30]. Additionally, chatbots often provide multiple responses rather than a single definitive answer, which can discourage self-diagnosis. Another challenge is determining the necessary information for accurate diagnosis. Some diseases, such as Hepatitis, which is blood-borne, can only be detected through blood tests, making self-diagnosis difficult for patients without the required materials and equipment. A major concern is the risk of private data exposure. As AI becomes more integrated into healthcare, a significant amount of patient information and health records are stored online. Due to the sensitive nature of these records, they are often targeted by hackers for fraud, who can remain undetected for extended periods [31].

Methodology

To reach a medical diagnosis, doctors usually ask numerous questions and may order lab tests if necessary. Symptoms play a crucial role in this process, as they can significantly narrow down the range of potential diseases. A patient’s medical history further refines the diagnostic process, making it more precise and aiding in identifying possible treatments. Lab tests are essential, providing valuable information that helps healthcare providers make informed decisions and manage patient care effectively. Additionally, reviewing a patient’s medication history can enhance the accuracy and safety of the diagnostic process.

In this study, to balance the need for a diverse range of cases with the constraints of time and expertise, three diseases were selected for evaluating the AAMD tools: Influenza, Clostridioides difficile (C. difficile), and Meningitis. These diseases vary significantly in prevalence: Influenza affects about 1 in 16 people in the US, C. difficile occurs in approximately 1 in 1,000 people, and Meningitis affects around 1 in 100,000 people.

The effectiveness of the four sets of parameters—symptoms, medical history, test results, and past medications—is measured by evaluating chatbots' diagnostic answers. More specifically, the effectiveness consists of two scores, namely an accuracy score and a suggestiveness score. The accuracy score indicates whether the disease was among the possible conditions identified by the chatbot, while the suggestiveness score reflects the proportion of accurate diagnoses among the suggestions provided.

Given the influence and popularity of various chatbots, we chose to use GPT-3.5 and GPT-4o for this research.

Table 1 provides the details of the cases used in this study. For each of the three target diseases, three cases were analyzed. The cases and their parameters were collected from literature and online sources [32-37]. It is important to note that the "Lab Tests" column includes only the definitive tests required to confirm the accuracy of the AAMD tools in diagnosing the disease.

Diseases	Cases	Symptoms	Medical History	Medications	Lab Tests
Influenza	Case I	fever, trouble breathing	no Vaccine at all, no chronic conditions, healthy	no anti-viral treatment	Influenza PCR
	Case II	fever, trouble breathing, upset stomach, chills, muscle aches	no vaccine at all, healthy	no anti-viral treatment	Influenza PCR
	Case III	fever, increased heart rate, low blood pressure	no vaccine, mild asthma	no anti-viral treatment	Influenza PCR
C. diff	Case I	constant diarrhea, stomach pain	visited sick grandmother with C. diff, healthy before	antibiotics for stye and parasite blastocystis hominis	EIA stool test
	Case II	no sleep, full body muscle spasms, hot sweats, cold sweats, migraine, constant diarrhea, stomach pain, bladder pain	get really ill if sick, sick for 8 weeks	antibiotics for sickness, 2nd antibiotics for sickness	EIA stool test
	Case III	sore throat, low body temperature, 97-102 temp	recently had colonoscopy	antibiotic for sickness	EIA stool test
Meningitis	Case I	throwing up, legs collapsed, in extreme pain	healthy before	no past medication	spinal tap
	Case II	fever, vomiting, body aches, lack of movement	healthy before	ibuprofen, Tylenol	spinal tap
	Case III	fever, headache, vision blurring, body aches, chills	healthy before	no past medication	spinal tap

Table 1 AAMD Test Cases

Results

The diagnostic parameters shown in Table 1 are transformed into questions and fed to both GPT-3.5 and GPT-4o. The resulting diagnostic answers are used to compile accuracy and suggestiveness scores, which are detailed in Table 2 for GPT-3.5 and Table 3 for GPT-4o. It's important to note that the symptom sets are chained with other parameter sets during the chat. Additionally, the accuracy rates for the definitive lab tests of the sample diseases are 100%, confirming that both GPT-3.5 and GPT-4o meet the basic sanity checks.

Table 2 indicates that combining symptoms, medical history, and medicine history yields the most accurate results for GPT-3.5, achieving an overall accuracy score of 66.7% and a suggestiveness score of 10.3%. Conversely, Table 3 shows that for GPT-4o, the best results are obtained with just symptoms and medicine history, with an overall accuracy score of 88.9% and a suggestiveness score of 8%. For both GPT-3.5 and GPT-4o, relying solely on symptoms results in lower accuracy, with scores of 44.4% for GPT-3.5 and 55.6% for GPT-4o.

Diseases	Cases	Symptoms		Symptoms + Medical History		Symptoms + Medications		Symptoms + Medical History + Medications		Lab Tests
		Acc.	Sug.	Acc.	Sug.	Acc.	Sug.	Acc.	Sug.	Acc.
Influenza	Case I	100%	1 of 7	100%	1 of 6	100%	1 of 8	100%	1 of 7	100%
	Case II	100%	1 of 8	100%	1 of 6	100%	1 of 5	100%	1 of 6	100%
	Case III	0%	0 of 6	100%	1 of 7	0%	0 of 7	100%	1 of 6	100%
C. diff	Case I	0%	0 of 8	0%	0 of 7	100%	1 of 7	100%	1 of 6	100%
	Case II	0%	0 of 7	0%	0 of 8	0%	0 of 7	0%	0 of 8	100%
	Case III	0%	0 of 6	0%	0 of 4	0%	0 of 6	0%	0 of 6	100%
Meningitis	Case I	0%	0 of 5	0%	0 of 7	0%	0 of 5	0%	0 of 6	100%
	Case II	100%	1 of 8	100%	1 of 7	100%	1 of 8	100%	1 of 6	100%
	Case III	100%	1 of 8	100%	1 of 10	100%	1 of 9	100%	1 of 7	100%

Table 2 Diagnostic Accuracy and Suggestiveness Scores for GPT-3.5

Diseases	Cases	Symptoms		Symptoms + Medical History		Symptoms + Medications		Symptoms + Medical History + Medications		Lab Tests
		Acc.	Sug.	Acc.	Sug.	Acc.	Sug.	Acc.	Sug.	Acc.
Influenza	Case I	100%	1 of 12	100%	1 of 10	100%	1 of 12	100%	1 of 14	100%
	Case II	100%	1 of 12	100%	1 of 10	100%	1 of 10	100%	1 of 8	100%
	Case III	0%	0 of 12	100%	1 of 10	100%	1 of 10	100%	1 of 10	100%
C. diff	Case I	0%	0 of 16	0%	0 of 9	100%	1 of 11	100%	1 of 10	100%
	Case II	0%	0 of 10	0%	0 of 8	100%	1 of 12	0%	0 of 18	100%
	Case III	0%	0 of 12	0%	0 of 12	0%	0 of 13	0%	0 of 10	100%
Meningitis	Case I	100%	1 of 10	100%	1 of 10	100%	1 of 10	100%	1 of 12	100%
	Case II	100%	1 of 15	100%	1 of 10	100%	1 of 10	100%	1 of 10	100%
	Case III	100%	1 of 16	100%	1 of 11	100%	1 of 12	100%	1 of 10	100%

Table 3 Diagnostic Accuracy and Suggestiveness Scores for GPT-4o

The suggestiveness scores are analyzed using standard statistical methods and are illustrated in Figures 1, 2, and 3 for Influenza, C. diff, and Meningitis, respectively.

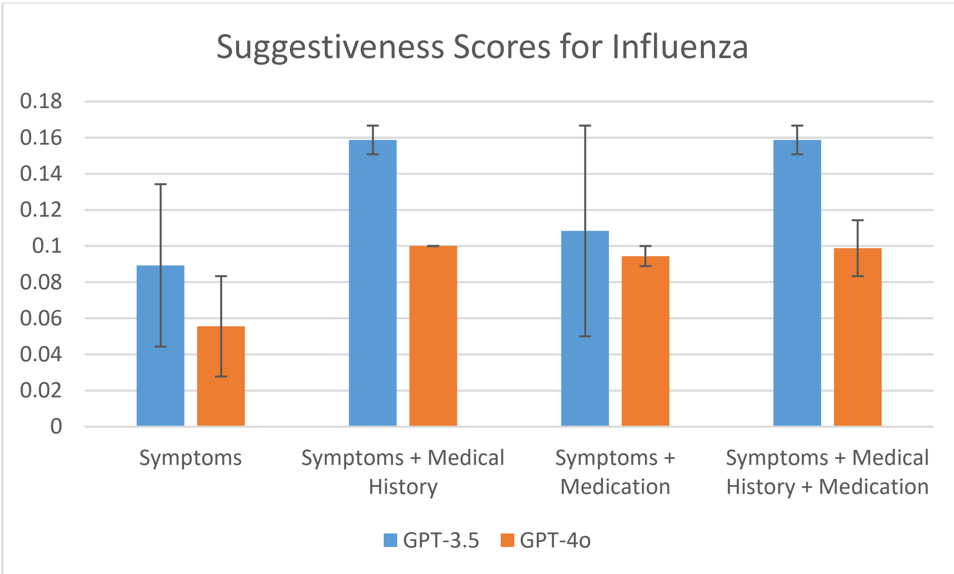

Figure 1 The Suggestiveness Scores for Influenza

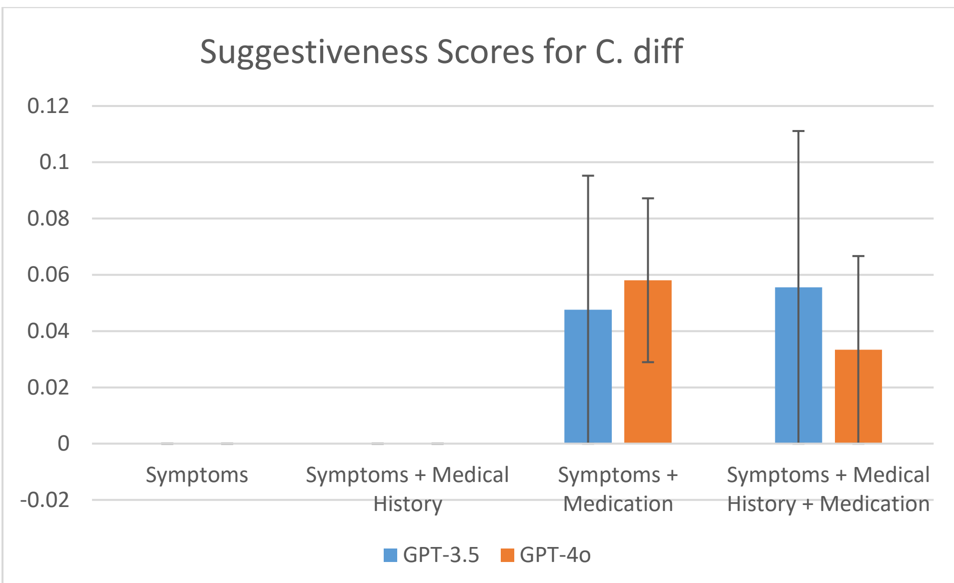

Figure 2 The Suggestiveness Scores for C. diff

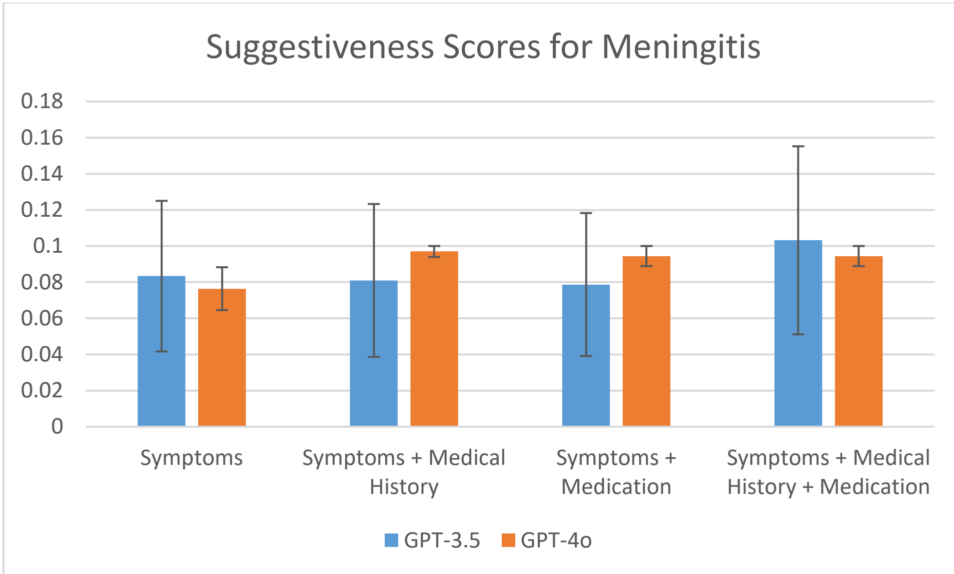

Figure 3 The Suggestiveness Scores for Meningitis

The experimental data reveals that diagnosing C. diff is difficult for both GPT-3.5 and GPT-4o, with GPT-3.5 achieving a combined accuracy score of 16.7% and GPT-4o achieving 25%. GPT-3.5 performs well in diagnosing Influenza, with an accuracy score of 83.3% and a suggestiveness score of 12.7%. In contrast, GPT-4o excels in diagnosing Meningitis, achieving a perfect accuracy score of 100% and a suggestiveness score of 8.82%.

The experimental data shows that GPT-4o is more likely to provide the correct diagnosis, with an accuracy score of 72.2%, compared to GPT-3.5's 55.6%. GPT-4o outperforms GPT-3.5 in accuracy across all three diseases tested. However, GPT-4o has a lower suggestiveness score than GPT-3.5, scoring 6.39% compared to 8.16%. On average, GPT-4o suggests 11.3 possible diagnoses, nearly twice the 6.8 options proposed by GPT-3.5.

To identify the optimal parameter set for AAMD tools, we conducted a two-way ANOVA (Analysis of Variance) test. This test examines how the mean suggestiveness score changes with different diagnostic parameter sets and sample diseases, and whether there is an interaction effect between these two independent variables.

Using data from Tables 2 and 3, Table 4 was created for the two-way ANOVA test, with parameter sets and sample diseases as the independent variables. The results, shown in Table 5, indicate a statistically significant difference in average suggestiveness scores based on the diagnostic parameter sets ($F(2)=18.92$, $p < 0.001$). However, the sample diseases and the interaction between these variables were not significant. Among the diagnostic parameter sets, the combination of symptoms and medical history proved to be the most effective.

	Parameter Set A	Parameter Set B	Parameter Set C	Parameter Set D
Influenza	0.14	0.17	0.13	0.14
	0.13	0.17	0.20	0.17
	0.00	0.14	0.00	0.17
	0.08	0.10	0.08	0.07
	0.08	0.10	0.10	0.13
	0.00	0.10	0.10	0.10
C. diff	0.00	0.00	0.14	0.17
	0.00	0.00	0.00	0.00
	0.00	0.00	0.00	0.00
	0.00	0.00	0.09	0.10
	0.00	0.00	0.08	0.00
	0.00	0.00	0.00	0.00
Meningitis	0.00	0.00	0.00	0.00
	0.13	0.14	0.13	0.17
	0.13	0.10	0.11	0.14
	0.10	0.10	0.10	0.08
	0.07	0.10	0.10	0.10
	0.06	0.09	0.08	0.10

Table 4 Suggestiveness Scores with Parameter Sets and Sample Diseases

ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Sample	0.092019	2	0.046009	18.92106	4.25E-07	3.150411
Columns	0.015451	3	0.00515	2.117974	0.107357	2.758078
Interaction	0.013254	6	0.002209	0.908466	0.495093	2.254053
Within	0.145899	60	0.002432			
Total	0.266623	71				

Table 5 Two-way ANOVA Test Results

Discussions

The study reveals that AAMD can assist patients in self-diagnosing to some extent. Both GPT-3.5 and GPT-4o perform well in diagnosing influenza and meningitis but have difficulty diagnosing C. diff. A notable concern is that GPT-4o's overall suggestiveness score is lower than that of GPT-3.5. The experiments show no correlation between the rarity of the disease and the effectiveness of AAMD diagnosis by either GPT-3.5 or GPT-4o. Both tools struggled with diagnosing C. diff, a disease of medium rarity, while GPT-4o achieved perfect accuracy in diagnosing meningitis, the rarest disease.

It is noteworthy that excessive information may negatively impact the accuracy of the diagnosis. GPT-4o yields better scores with only symptoms and medicine history than with a full set of parameters including symptoms, medical history, and medicine history.

As previously noted, GPT-4o is more likely to include the correct diagnosis in its list but provides significantly more options than GPT-3.5. To optimize GPT-4o for accuracy while reducing the number of possible diagnoses, the question structure should be adjusted. The current question was designed for GPT-3.5, which is more open-ended, whereas GPT-4o is limited in response capacity. By removing the "List possible diseases" section from the query, the number of diagnoses generated by GPT-4o decreases. However, omitting this section from GPT-3.5 would significantly reduce its diagnostic effectiveness, as it tends to avoid vague responses and does not directly suggest possible diagnoses.

While using GPT-3.5 and GPT-4o, we observed a particularly intriguing phenomenon – GPT seems to be language-dependent. If the same question is asked in different languages, the answers from GPT can be vastly different. In future research, it might be a good idea to investigate if the chat language is a contributing factor in AAMD.

In conclusion, this study evaluates the effectiveness of common parameters provided to generative AI for medical diagnostics, including symptoms, medical history, medical tests, and medications. Statistical analysis of accuracy and suggestiveness scores indicates that patient symptoms and medical history are the most crucial parameters for AI-assisted medical diagnostics. This study can be expanded to include all diagnosable diseases and other AI tools. Overall, AAMD can enhance the efficiency of the diagnostic process for both patients and doctors, potentially saving many lives and reducing costs.

References

[1] Biswas, S. S. (2023). Potential use of Chat GPT in global warming. Annals of Biomedical Engineering, 51(6), 1126–1127. https://doi.org/10.1007/s10439-023-03171-8

[2] Biswas, S. S. (2023b). Role of Chat GPT in Public Health. Annals of Biomedical Engineering, 51(5), 868–869. https://doi.org/10.1007/s10439-023-03172-7

[3] McGee, R. W. (2023). Annie Chan: Three Short Stories Written with Chat GPT. Available at SSRN 4359403.

[4] McGee, R. W. (2023). Is Chat GPT biased against Conservatives? An Empirical study. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.4359405

[5] Mathew, A. . (2023). Is Artificial Intelligence a World Changer? A Case Study of OpenAI's Chat GPT. Recent Progress in Science and Technology Vol. 5, 35–42. https://doi.org/10.9734/bpi/rpst/v5/1 8240D

[6] Borji, A. (2023, February 6). A categorical archive of ChatGPT failures. arXiv.org. https://arxiv.org/abs/ 2302.03494

[7] Alkaissi H, McFarlane S I (February 19, 2023) Artificial Hallucinations in ChatGPT: Implications in Scientific Writing. Cureus 15(2): e35179. doi:10.7759/cureus.35179

[8] Frieder, S., Pinchetti, L., Griffiths, R., Salvatori, T., Lukasiewicz, T., Petersen, P., & Berner, J. (2023, December 15). Mathematical capabilities of ChatGPT. https://proceedings.neurips.cc/paper_files /paper/2023/hash/58168e8a92994655d6da3939e7cc0918-Abstract-Datasets_and_Benchmarks.html

[9] Ayub, H. (2024, May 14). GPT-4O: Successor of GPT-4? - Hamid Ayub - Medium. Medium. https://ham idayub.medium.com/gpt-4o-successor-of-gpt-4-8207acf9104e#:~:text=Parameter%20Count,billion%20parameters%20of%20GPT%2D4.

[10] Ray, P. P. (2023). ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. Internet of Things and Cyber-Physical Systems, 3, 121–154. https://doi.org/10.1016/j.iotcps.2023.04.003

[11] Paulus, N. (2024, May 22). How many Americans are uninsured? MoneyGeek.com. https://www.mo neygeek.com/insurance/health/analysis/americans-without-coverage/

[12] LaPook, J. (2024, July 7). Mental health chatbots powered by artificial intelligence developed as a therapy support tool. CBS News. https://www.cbsnews.com/news/mental-health-chatbots-powe red-by-artificial-intelligence-providing-support-60-minutes-transcript/

[13] Pinto-Coelho, L. (2023). How Artificial intelligence is shaping medical Imaging Technology: A survey of Innovations and applications. Bioengineering, 10(12), 1435. https://doi.org/10.3390/bioengin eering10121435

[14] Emerging capabilities in the science of Artificial Intelligence - Giving to Mayo Clinic. (2024, February 20). https://www.mayoclinic.org/giving-to-mayo-clinic/our-priorities/artificial-intelligence

[15] Noble.Dana. (2024, April 17). AI in healthcare: The future of patient care and health management. Mayo Clinic Press. https://mcpress.mayoclinic.org/healthy-aging/ai-in-healthcare-the-future-of-p atient-care-and-health-management/

[16] Massachusetts Institute of Technology. (2024, March 27). Mayo Clinic's Healthy Model for AI success | Thomas H. Davenport and Randy Bean | MIT Sloan Management Review. MIT Sloan Management Review. https://sloanreview.mit.edu/article/mayo-clinics-healthy-model-for-ai-succ ess/

[17] Clinic, C. (2024, April 30). AI Center to advance care for congenital heart Disease. Cleveland Clinic. ht tps://consultqd.clevelandclinic.org/ai-center-to-advance-care-for-congenital-heart-disease

[18] Cleveland Clinic Children's Center for Artificial Intelligence (C4AI) | Cleveland Clinic Children's. (n.d.). Cleveland Clinic. https://my.clevelandclinic.org/pediatrics/medical-professionals/artificial-intellig ence

[19] AI enables faster, more precise image registration for medical image. (2024, July 29). Mass General Advances in Motion. https://advances.massgeneral.org/radiology/q-a.aspx?id=1028

[20] Medical artificial intelligence with a purpose. (2023, June 21). https://carey.jhu.edu/articles/researc h/medical-artificial-intelligence-purpose#:~:text=Johns%20Hopkins%20is%20a%20pioneer,in%20a%20pediatric%20care%20setting.

[21] Jacobs, D. (2020, February 24). Practical uses for artificial intelligence in health care. Johns Hopkins Medicine. https://www.hopkinsmedicine.org/news/articles/2020/02/practical-uses-for-artificial-intelligence-in-health-care

[22] Artificial intelligence and the future of health care. (2023b, May 25). UCLA Health. https://www.ucla health.org/news/article/artificial-intelligence-and-future-health-care

[23] Sennaar, K. (2020b, March 24). How America's 5 Top Hospitals are Using Machine Learning Today. Emerj Artificial Intelligence Research. https://emerj.com/ai-sector-overviews/top-5-hospitals-usi ng-machine-learning/

[24] Gervais, G. (2023, October 12). Carta Healthcare survey results indicate that education around AI may improve consumer trust. Business Wire. https://www.businesswire.com/news/home/20231 012647360/en/Carta-Healthcare-survey-results-indicate-that-education-around-AI-may-improve-consumer-trust

[25] Al-Medfa, M. K., Al-Ansari, A. M., Darwish, A. H., Qreeballa, T. A., & Jahrami, H. (2023). Physicians' attitudes and knowledge toward artificial intelligence in medicine: Benefits and drawbacks. Heliyon, 9(4), e14744. https://doi.org/10.1016/j.heliyon.2023.e14744

[26] Coppola, F., Faggioni, L., Regge, D. et al. Artificial intelligence: radiologists' expectations and opinions gleaned from a nationwide online survey. Radiol med 126, 63–71 (2021). https://doi.org/10.100 7/s11547-020-01205-y

[27] American Medical Association & American Medical Association. (2024, January 12). Big majority of doctors see upsides to using health care AI. American Medical Association. https://www.ama-ass n.org/practice-management/digital/big-majority-doctors-see-upsides-using-health-care-

ai#:~:text=Physicians%20have%20guarded%20enthusiasm%20for,physician%20relationship%20and%20patient%20privacy.

[28] Davenport, T., & Kalakota, R. (2019). The potential for artificial intelligence in healthcare. Future Healthcare Journal, 6(2), 94–98. https://doi.org/10.7861/futurehosp.6-2-94

[29] Kolata, G. (2023, July 23). A Mystery in the E.R.? Ask Dr. Chatbot for a Diagnosis. The New York Times. https://www.nytimes.com/2023/07/22/health/chatbot-medical-mystery-diagnosis.html

[30] Shieh, A., Tran, B., He, G., Kumar, M., Freed, J. A., & Majety, P. (2024). Assessing ChatGPT 4.0's test performance and clinical diagnostic accuracy on USMLE STEP 2 CK and clinical case reports. Scientific Reports, 14(1). https://doi.org/10.1038/s41598-024-58760-x

[31] Khan, B., Fatima, H., Qureshi, A., Kumar, S., Hanan, A., Hussain, J., & Abdullah, S. (2023). Drawbacks of artificial intelligence and their potential solutions in the healthcare sector. Deleted Journal, 1(2), 731–738. https://doi.org/10.1007/s44174-023-00063-2

[32] Meyer, A. N., Giardina, T. D., Khawaja, L., & Singh, H. (2021). Patient and clinician experiences of uncertainty in the diagnostic process: Current understanding and future directions. Patient Education and Counseling, 104(11), 2606–2615. https://doi.org/10.1016/j.pec.2021.07.028

[33] Nichol, J. R., Sundjaja, J. H., & Nelson, G. (2024, April 30). Medical history. StatPearls - NCBI Bookshelf. https://www.ncbi.nlm.nih.gov/books/NBK534249/#:~:text=Both%20aspects%20of%20the%20history,and%20determining%20appropriate%20imaging%20modalities.

[34] Francis, M., Deep, L., Schneider, C. R., Moles, R. J., Patanwala, A. E., L, L., DO, Levy, R., Soo, G., Burke, R., & Penm, J. (2022). Accuracy of best possible medication histories by pharmacy students: an observational study. International Journal of Clinical Pharmacy, 45(2), 414–420. https://doi.org/10.1007/s11096-022-01516-2

[35] Weekly U.S. influenza surveillance report. (2024, August 16). Centers for Disease Control and Prevention. https://www.cdc.gov/flu/weekly/index.htm#:~:text=CDC%20estimates%20that%20there%20have,important%20for%20higher%20risk%20patients.

[36] About C. diff. (2024, March 6). C. Diff (Clostridioides Difficile). https://www.cdc.gov/c-diff/about/index.html#:~:text=%2D%20C.,the%20subsequent%202%2D8%20weeks.

[37] Watkins Health Services | The University of Kansas. (n.d.). Watkins Health Services. https://studenthealth.ku.edu/meningitis#:~:text=Meningococcal%20meningitis%20is%20a%20rare,approximately%20150%20to%20300%20deaths.